Applied Water Science (2018) 8:167
https://doi.org/10.1007/s13201-018-0794-7

ORIGINAL ARTICLE

Evaluation of biological wastewater treatment process using Mahalanobis distances in original and principal component space: a case study

Petr Praus[1]

Received: 6 February 2018 / Accepted: 31 August 2018

Abstract

The evaluation of wastewaters treated by biological wastewater treatment plant was performed using multivariate analysis. The samples taken during 1 year were characterized by the Mahalanobis distances (MDs) calculated from 11 original parameters and 4 principal components extracted by principal component analysis. The principal components were interpreted using Ward's hierarchical clustering analysis and factor analysis. Statistical processing of the samples by means of the MDs calculated in the original and PC space was found to be complementary. Since MDs were not normally distributed, the statistical analysis of their log-transformed values (logMDs) was preferred to common Hotelling's or Chi-squared statistics of MD^2 ones. The outliers were confirmed by Ward's method and by inspection of their chemical composition. In contrast to complexity and different magnitudes of the original wastewater parameters, the logMD charts provided a simple and effective tool for the evaluation of biological wastewater treatment process.

Keywords Statistical evaluation · Biological wastewater treatment · Mahalanobis distance · Principal component analysis

Introduction

Quality monitoring of processes has a long tradition in many technical and economic fields (Montgomery 1980, 1996). The main reason is detection of process shifts and out-of-control events. Some applications of statistical control in non-industrial processes and in environment monitoring were reviewed by Cobert and Pan (2002). Shewhart's control charts (Shewhart 1939) of selected individual variables were used for the evaluation of sewage treatment stations (Berthoex et al. 1978; Orssatto et al. 2014) and river water quality (Iglesias et al. 2016). The use of a cumulative sum (CUSUM) method for monitoring of water quality in storages was described by Mac Nally and Hart (1997) and also discussed, for example, by Cobert and Pan (2002).

Water composition can be characterized by many chemical and physical variables (parameters), and evaluation of water quality is a multidimensional problem. Univariate control charts of individual parameters can lead to erroneous conclusions. They may incorrectly identify out-of-control situations (Montgomery 2009). Water quality parameters are mutually correlated, of different extent and out of normal distribution (Vega et al. 1998). Hotelling's control chart (Hotelling 1947) as well as other multivariate charts like CUSUM and exponentially weighted moving average (EWMA) charts was theoretically described in some review papers, e.g. MacGregor and Kourti (1995). For example, the application of Hotelling's control chart in the monitoring of BWWT process was referred by Capilla (2009).

The Mahalanobis distance allows computing the distance between two objects in an n-dimensional space (Mahalanobis 1936). PCA is used to reduce the dimensionality of original data to several principal components, to help us understand relationships among original variables and also to form clusters of similar objects. The visualization of the objects in two- or three-dimensional systems is also a big advantage of this method.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13201-018-0794-7) contains supplementary material, which is available to authorized users.

✉ Petr Praus
petr.praus@vsb.cz

[1] Institute of Environmental Technology, VŠB-Technical University of Ostrava, 17. listopadu15, 70830 Ostrava-Poruba, Czech Republic

Published online: 21 September 2018

The aim of this paper is to demonstrate the utilization of the Mahalanobis distance for the characterization of treated wastewater composition. The samples were characterized by 11 parameters, such as biochemical oxygen demand after 5 days (BOD), chemical oxygen demand (COD), total phosphorus (TP), total nitrogen (TN), total suspended solids (TSS), total dissolved solids (TDS), pH, ammonium nitrate, phosphate and cyanide. The MDs were calculated in the original data space and reduced PC space and evaluated by one-dimensional charts. The samples were also clustered by PCA and Ward's clustering method into groups of different compositions.

Methods

Data collection

The compositions of treated wastewater were characterized by 11 parameters such as BOD, COD, TP, TN, TSS, TDS, pH, ammonium, nitrate, phosphate and cyanide. Water analyses, including sampling and preservation, were carried out according to ISO and EN standard procedures: EN 1899-1: 1998 (BOD), ISO 6060: 1989 (COD), EN ISO 6878: 2004 (TP and phosphate), EN 25663:1993 (TN), EN 872:1996 (TSS and TDS), ISO 10523:2008 (pH), ISO 7150-1:1984 (ammonium), ISO 7890-3:1988 (nitrate) and ISO 6703-1:1984 (cyanide).

Spectrophotometric determination of ammonium, nitrate, phosphate and cyanide was performed using an UV–VIS spectrometer DR 4200 (HACH). TDS and TSS were determined gravimetrically after sample filtration through 0.85-µm membrane filters. pH was determined with a device pH 197 (WTW).

The samples were collected and measured daily during 1 year at an outlet of BWWTP designed for the capacity of 638,850 population equivalents for the treatment of municipal and industrial wastewater produced on the territory of an industrial city with the current population of about 300,000. Industrial water was produced mostly in chemical, metallurgical and coke-making processes. The basic statistics of the analysed samples are summarized in Table 1.

Principal component analysis

The main objective of PCA is a search for new latent (hidden) variables of n samples which are orthogonal (not correlated) to each other (Jolliffe 2002). Each latent variable–principal component is a linear combination of p variables x_i and describes a different source of total variation

$$t_{im} = x_{i1}w_{1m} + x_{i2}w_{2m} + \cdots x_{ip}\,w_{pm} \quad (1)$$

Table 1 Basic statistics of treated wastewater samples ($n = 334$)

	BOD (mg/l)	COD (mg/l)	CN^- (mg/l)	NH_4^+ (mg/l)	NO_3^- (mg/l)	pH	PO_4^{3-} (mg/l)	TDS (mg/l)	TN (mg/l)	TP (mg/l)	TSS (mg/l)
Aver.	4.9	33.9	0.106	0.63	50.8	7.9	1.31	759	17.1	0.62	7.2
Var.	4.2	43.2	0.002	0.82	130	0.02	0.58	12,014	12.1	0.07	9.0
SD	2.1	6.6	0.043	0.90	11.4	0.2	0.76	110	3.5	0.26	3.0
Min.	1.1	15.8	0.016	0.05	14.8	7.2	0.026	342	6.9	0.12	2.0
Max.	13.8	75.2	0.267	6.05	94.6	8.3	3.62	972	47.5	1.64	22.0
Median	4.6	33.9	0.100	0.23	51.0	7.9	1.15	792	17.2	0.58	7.0
Skew.	1.35	0.66	0.762	2.92	0.383	−0.28	0.822	− 1.17	1.75	0.862	1.33
Kurt.	5.08	7.04	3.49	12.7	3.85	3.79	3.13	4.18	19.8	3.43	6.71

where t_{im} is the score of the i-th object in the m-th component. The component loadings w_{im} are contribution measures of a particular variable to the PCs. The variability of the PCs is given by the corresponding eigenvalues λ_m, where $m = 1, 2, \ldots p$, which are ordered as $\lambda_1 > \lambda_2 > \cdots \lambda_p$, where each eigenvalue is a variance of the corresponding m-th component. PCA can be performed by eigenvalue decomposition of a correlation (or covariance) matrix or by singular value decomposition of a data matrix (Praus 2005a, b).

Mahalanobis distance

Mahalanobis distance of a variable x_i can be calculated as

$$\mathrm{MD}_i = \sqrt{(x_i - \mu)^{\mathrm{T}} C^{-1} (x_i - \mu)} \quad (2)$$

where μ is the mean of n variables x_i and $\boldsymbol{C}$ is the covariance matrix. If the covariance matrix is an identity matrix, the Mahalanobis distance reduces itself to the Euclidean distance $\mathrm{ED}_i = \sqrt{(x_i - \mu)^{\mathrm{T}} (x_i - \mu)}$. Also, in special cases, where parameters are uncorrelated and variances in all directions are the same, the Mahalanobis distance becomes equivalent to the Euclidean distance.

Factor analysis

In factor analysis each variable can be expressed as a linear combination of latent common factors and a single specific factor as follows

$$x_{im} = F_{i1} f_{1m} + F_{i2} f_{2m} + \cdots + F_{ip} f_{pm} + e_{im} \quad (3)$$

where F_{ip} and e_{im} are the common and specific (error) factors, respectively, f_{im} are the factor loadings. FA separates the correlation (covariance) matrix into two matrices: a common factor matrix and a specific factor matrix. The main difference between PCA and FA is that while PCA concerns the total variation as expressed in the correlation matrix, FA concerns a correlation in the common factor portion. The number of factors must be known before FA is performed. The methods of factor computations including the detailed explanation of FA are described in the literature, e.g. Malinowski (1991).

Cluster analysis

Cluster analysis encompasses a lot of different methods that arrange objects into groups according to their similarity. This exploratory method is used to discover a data structure not only among observations, but also among variables, arranged into dendrograms. Utilized methods, algorithms and similarity/dissimilarity measures are described elsewhere in the literature (e.g. Everitt 2001). In this study, common Ward's hierarchical clustering method (Ward 1963) was used for clustering of water samples.

Statistic calculations

An original data matrix of 345 wastewater samples was set up and processed in MS Excel. Missing samples were distributed randomly during a year; they were not collected during weekends and days off. PCA, FA, CA and other statistical calculations were performed using software packages STATGRAPHICS Plus 5.0 (Statistical Graphic Corp.), QC.Expert (Trilobyte, Czech Republic) and XLSTAT 2017 (Addinsoft). Before the multivariate analysis, 11 outliers with extremely high magnitudes of COD, BOD and ammonium were identified using the box-and-whisker plots and excluded from the dataset, and the remaining 334 samples were statistically analysed like that. The data were always standardized in order for us to avoid misclassifications arising from different orders of magnitude of variables. For this purpose, the data were mean (μ)-centred and scaled by standard deviations (σ) as $(x-\mu)/\sigma$.

Results and discussion

Mahalanobis distances in original data space

Mahalanobis distance is a very useful measure in multivariate analysis because it simply shows the distance of an object in n-dimensional space from the centre of a dataset. That way each object (sample) can be visualized as a point in one diagram instead of several diagrams being used and constructed for each variable. In this study, the Mahalanobis distances were calculated for treated water samples taken daily at BWWTP and plotted in charts, which showed their time changes during the year. The samples were labelled with ordinal numbers of their sampling; therefore, the sample coordinate corresponded to time one.

The samples were characterized by 11 variables, from which the Mahalanobis distances were calculated according to Eq. (2). Their distribution was far from normality, and therefore, common Hotelling's and Chi-squared statistics were not used. The logarithms of MDs (logMDs) were calculated in order for us to obtain their normal distribution. It was confirmed by the Kolmogorov–Smirnov test ($p = 0.325$) and a probability–probability plot. The logMD magnitudes of all samples are shown in Fig. 1.

The average μ and standard deviation σ of the logMD magnitudes were calculated as $\mu = 0.464$ and $\sigma = 0.143$, respectively, and the 2σ and 3σ values were calculated as 0.286 and 0.429, respectively, as shown in Fig. 1. The $\mu + 2\sigma$ and $\mu + 3\sigma$ limits can play a role of upper warning (UWL) and upper control (UCL) limits used in the well-known

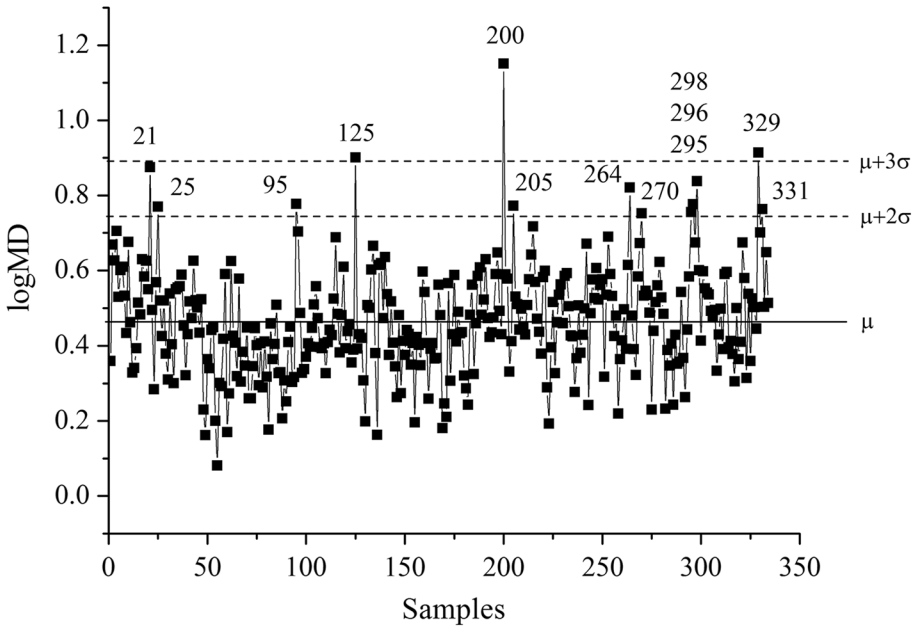

Fig. 1 Chart of logMDs calculated from 11 original variables

Shewhart's charts. The samples outside these limits indicate that the process is not running as consistently as possible. The samples above $\mu+3\sigma$ and between the $\mu+2\sigma$ and $\mu+3\sigma$ limits are considered as outliers and should be of interest to BWWTP operators. They are usually explained by investigation of individual parameters of the outlaying samples.

In order for us to get insight into BWWT process, the data dimensionality was reduced by PCA and the MDs were calculated in a new PC space.

Sample evaluation in PC space

Principal component analysis

PCA was performed by decomposition of the correlation matrix composed of the above given samples in order for us to calculate eigenvalues and corresponding eigenvectors. Bartlett's sphericity test rejected the null hypothesis that there was no correlation significantly different from 0 ($p<0.0001$). Based on the PCA results, first, relationships between the parameters were discussed and, second, the samples were characterized by a few PCs from which the MDs were calculated.

There is no universal rule for how to estimate the number of principal components. According to the magnitude of eigenvalue, which should be equal to or higher than 1 (Kaiser 1960), 4 principal components explaining about 72% of the total data variance were selected (Table 2). That way the original data dimensionality was reduced from 11 to 4 and components with lower variability were neglected. The next step of PCA often includes interpretation of the principal components.

Table 2 Principal component analysis of standardized data

Component number	Eigenvalue	Variance (%)	Cumulative variance (%)
1	3.090	28.1	28.1
2	1.759	16.0	44.1
3	1.639	14.9	69.0
4	1.378	12.5	71.5
5	0.8600	7.8	79.3
6	0.6457	5.9	85.2
7	0.5472	5.0	90.2
8	0.3789	3.4	93.6
9	0.3748	3.4	97.0
10	0.2379	2.2	99.2
11	3.090	0.8	100

Interpretation of principal components

Interpretation of PCs is necessary and useful for our understanding of the data structure. The component loadings can elucidate relationships among original variables. Component loadings of the PCs are summarized in Table 3.

The first principal component (PC1) was saturated mainly with organic and inorganic compounds present in municipal wastewater, which can be characterized by common parameters, such as BOD, COD, TN, ammonium, TP, TSS and TDS. The second principal component (PC2) was influenced by nitrate and also BOD and TSS. Nitrate indicates nitrification processes taking part in activated sludge. TSS characterizes solid particles, e.g. of activated sludge, released into treated water.

Table 3 Loadings of main 4 principal components

Parameters	PC1	PC2	PC3	PC4
NH_4^+	0.437	0.323	0.091	0.029
BOD	0.532	0.642	−0.019	0.146
COD	0.696	0.274	−0.055	0.070
NO_3^-	0.446	−0.595	−0.514	−0.002
PO_4^{3-}	0.492	−0.333	0.689	−0.348
CN^-	0.166	−0.115	0.444	0.681
TN	0.669	−0.389	−0.422	0.133
TSS	0.547	0.638	−0.220	−0.044
TP	0.669	−0.154	0.575	−0.354
pH	−0.159	−0.046	0.331	0.724
TDS	0.666	−0.328	−0.158	0.312

Table 4 Loadings of main 4 factors after Varimax rotation

Parameters	Factor 1	Factor 2	Factor 3	Factor 4
NH_4^+	0.5230	0.0312	0.1674	0.0446
BOD	0.8432	−0.0199	−0.0084	0.0696
COD	0.6678	0.2970	0.1808	0.0206
NO_3^-	−0.1244	0.8768	0.0109	−0.1800
PO_4^{3-}	−0.0111	0.0693	0.9703	0.0471
CN^-	0.0613	0.0752	0.1501	0.8180
TN	0.1822	0.8690	0.0673	−0.0338
TSS	0.8442	0.0560	−0.0662	−0.1908
TP	0.2449	0.1267	0.9225	−0.0237
pH	−0.0860	−0.0928	−0.1228	0.7942
TDS	0.2274	0.7325	0.1644	0.2401

The third principal component (PC3) was mostly influenced by phosphate and TP, which of course well correlate with each other. Unlike PC1, the relatively high negative loading of nitrate was likely caused by higher variance of nitrate in comparison with phosphate, resp. TP (Table 1). Correlations between NO_3^- and PO_4^{3-}, resp. TP, were very low (Table S1, Supplementary materials). The fourth principal component (PC4) was mainly saturated with pH and cyanides. Cyanides were a part of alkaline effluent of a coking plant entering BWWTP. A special inlet for these waters was built directly in an activated sludge tank. Therefore, high positive correlation of both parameters is rational.

The relationships among the parameters were verified by factor analysis after Varimax rotation (Table 4). Unlike PCA, FA is also used to interpret revealed main factors and relationships between variables. In factor 1, NH_4^+, BOD, COD and TSS were well correlated, which was due to the existence of organic compounds in treated wastewater. In factor 2, positive correlation between TN and NO_3^- confirmed that most of the nitrogen compounds were oxidized into nitrate during the nitrification. Their correlation with TDS showed that nitrate was the prevailing anion in treated water. On the whole, PC1 and PC2 as well as factor 1 and factor 2 described the performance of the activated sludge treatment process. Relationships of TP and PO_4^{3-} and pH and CN^- demonstrated in factors 3 and 4, respectively, are in agreement with the compositions of PC3 and PC4, respectively.

For the interpretation of the PCs, Ward's hierarchical clustering method was used as well (Figure S1). It confirmed the compositions of PC3 and PC4 as well as factors 3 and 4. Some differences were found in the first and second PCs and factors as a result of close relationships between these parameters and fundamentals of the statistical methods used.

Normal distribution of the PCs was confirmed by the Kolmogorov–Smirnov test ($p = 0.808$, 0.229, 0.252 and 0.481, respectively). The average μ and standard deviation σ of each PC can be calculated as well as their statistical limits 2σ and 3σ. Thus, the $\mu \pm 2\sigma$ (UWL) and $\mu \pm 3\sigma$ (UCL) limits can be used for the construction of the same control charts shown in Fig. 1. It was not described here.

Mahalanobis distances in PC space

After PCA, each water sample was considered a point in a new four-dimensional PC space. Their MDs were calculated using the correlation matrix of the 4 PC scores according to Eq. (2) (MacGregor and Kourti 1995; Maesschalck et al. 2000).

The MDs were far from normal distribution, and therefore, their common logarithms (logMD) were calculated in order for us to obtain their normal distribution, which was confirmed by the Kolmogorov–Smirnov test ($p = 0.412$). For the statistical evaluation, the logMDs are plotted for all samples in Fig. 2. The average μ and standard deviation σ of the logMDs were calculated as $\mu = 0.238$ and $\sigma = 0.168$, and the statistical 2σ and 3σ limits were 0.336 and 0.504, respectively. The outlaying samples between the $\mu + 2\sigma$ and $\mu + 3\sigma$ limits are marked with their numbers. In order for us to explain the outliers, the charts of PCs scores were constructed as shown in Fig. 3.

For example, sample 21 was indicated due to high PC1 and low PC2 scores, which was caused by high concentrations of BOD, COD and ammonium and low concentration of nitrate. Sample 25 was different due to high concentrations of nitrate and total nitrogen and low concentration of BOD, which has a negative loading in PC2. The extreme sample 96 was explained by low scores of PC1, PC3 and PC4 due to low content of organic compounds, phosphorus-based compounds and TDS. Sample 188 was indicated as an outlier due to low content of ammonium, organic compounds and TDS. On the other hand, sample 200 had high content of nitrate and high pH, which shows that the

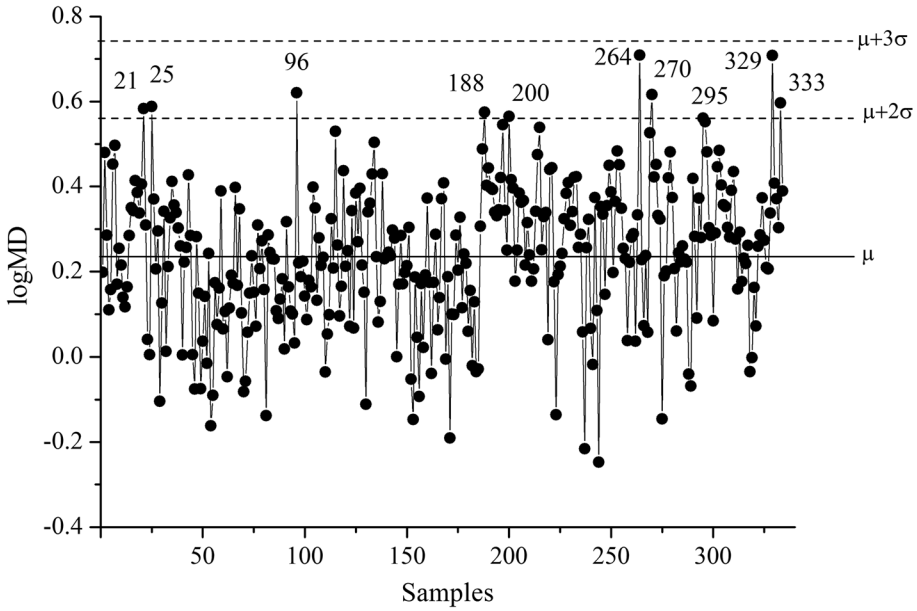

Fig. 2 Chart of logMDs calculated from PCs scores

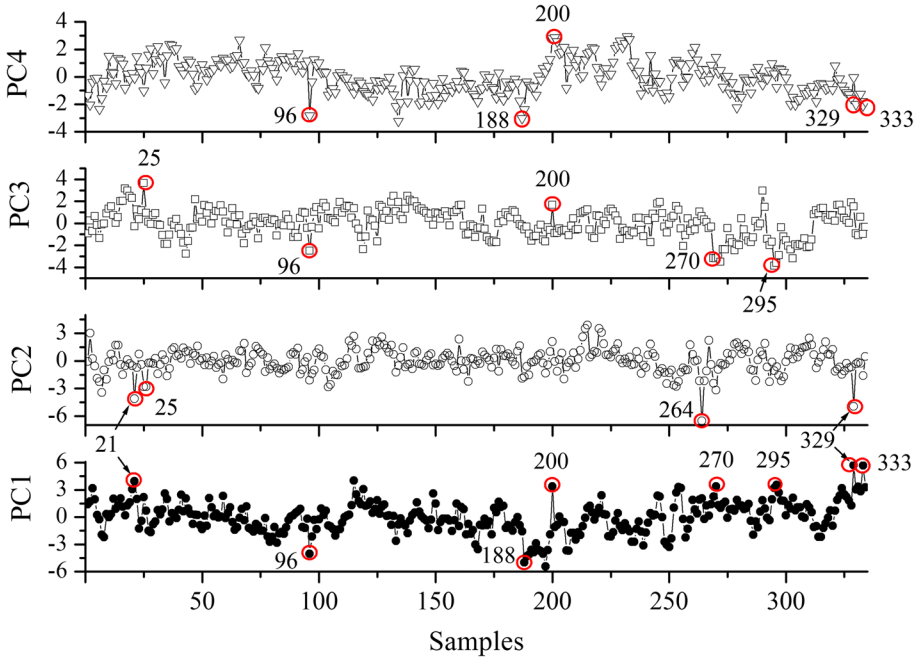

Fig. 3 Charts of 4 main principal components. Outliers are marked by circles

denitrification process was inhibited by high concentration of hydroxide ions, for example, as follows

$$5\,CH_3OH + 6\,NO_3^- \rightarrow 5\,CO_2 + 7\,H_2O + 6\,OH^- + 3\,N_2 \quad (4)$$

The high pH was likely caused by the coking plant effluent although the cyanide concentration was within the current magnitudes. In addition, this outlier also indicates a typing or laboratory mistake because the concentrations of nitrate and TN cannot be so similar.

Sample 264 was influenced by the low scores of PC2, that is, by the high content of BOD and TSS. Samples 270 and 295 were different due to the concentrations of phosphate and cyanide anions. In these samples, BOD and ammonium magnitudes were high, respectively. Samples 329 and 333 contained the high concentrations of organic compounds and high concentrations of TP and also TDS in sample 333. The concentrations of nitrate were also high and thus compensated the negative loadings of phosphate and TP in PC3.

All the outliers mentioned above can indicate unusual situations that occurred during the treatment process. The outliers were further verified by samples clustering and comparison with the results of chemical analyses.

Verification of MDs-based evaluation

Ward's cluster analysis

The samples in the PC space were clustered by Ward's clustering method and grouped into 3 main clusters shown in dendrogram. Figure 4 shows the left, central and right clusters containing samples 107, 54 and 173, respectively. The cluster centroids are summarized in Table 5.

These 3 clusters are also shown in the PC scatter plot in Fig. 5. Two dashed lines approximately separate the samples into 3 clusters according to the Ward's dendrogram. The samples identified in Fig. 3 are marked by circles. The samples in the central cluster had the negative scores of PC1 and PC2, that is, the lowest concentrations of parameters indicating organic contamination. As given above, PC1 and PC2 indicated the efficiency of the activation treatment of wastewater, and therefore, these samples correspond to the cleanest treated water as a result of the most effective treatment process and/or relatively little polluted incoming raw wastewater. On the contrary, the samples in the left cluster and especially in the first quadrant of the scatter plot (Fig. 5) correspond to cases in which the treated water was of worse quality than usual.

For comparison, the samples were clustered in the original data space into 4 clusters containing 90, 184, 38 and 22 samples (Figure S2). The cluster centroids are given in Table S2. The differences between the centroids were smaller than between those in the PC space. It confirms the ability of PCA to extract important information and remove noise from the data.

Table 5 Cluster centroids in PC space

Cluster	PC1	PC2	PC3	PC4
Left	1.717	−0.216	−0.597	−0.378
Central	−2.230	−1.034	−0.363	−0.861
Right	−0.366	0.456	0.482	0.502

Assessment of outlier compositions

Compositions of the 16 outlaying samples identified in the original and PC space are summarized in Table 6. They can be explained by the presence of at least two parameters whose magnitudes were different from the others. There were 13 samples and 10 samples identified in the original and PC space, respectively. Seven samples of all evaluated ones were identified by both procedures.

Different results obtained in the original and PC spaces can be explained by different complexity of data. The reduction of data dimensionality from 11 to 4 also led to the reduction of information corresponding to 28% of total variance. In this way, noise and also small part of information were removed. Therefore, the statistical evaluation by the Mahalanobis distances in reduced as well as original space is complementary and useful.

Characterization of water samples by the Mahalanobis distances calculated from original parameters and principal

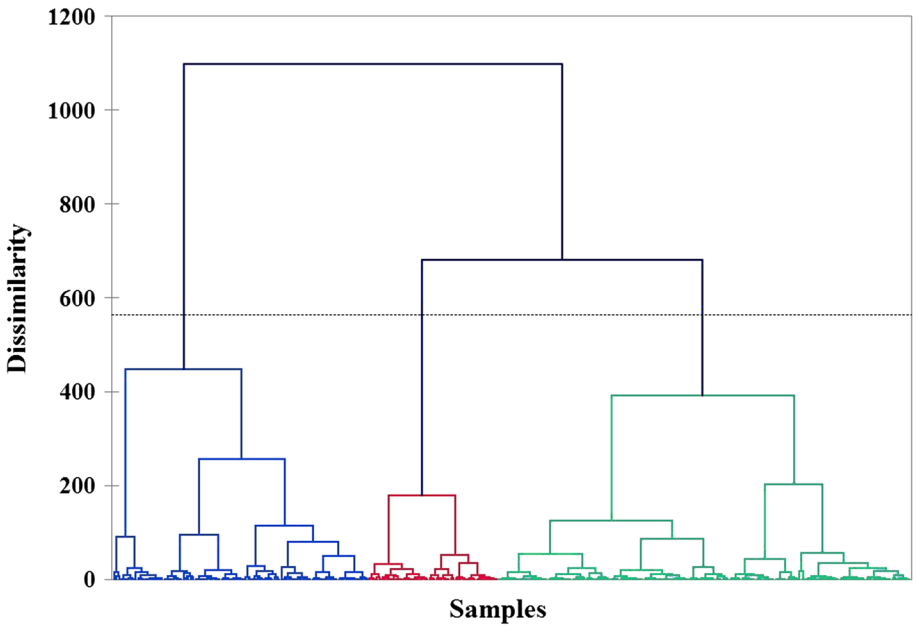

Fig. 4 Ward's dendrogram of treated wastewater samples in their PC space

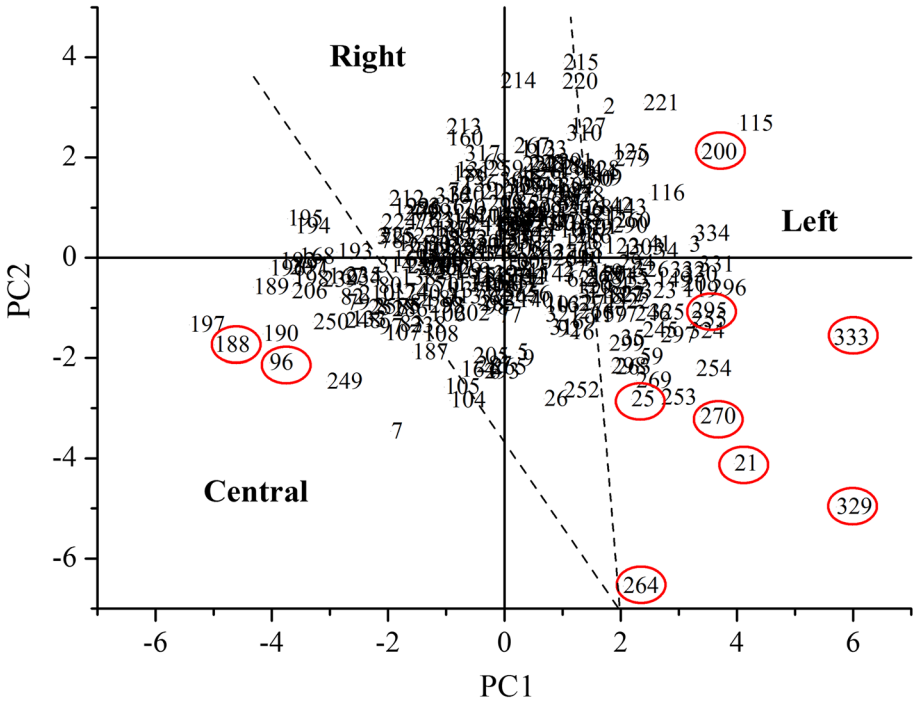

Fig. 5 PCs scatter plot of treated wastewater samples

components scores is proposed for the evaluation of water samples collected during long periods as well as for daily statistical control of BWWTPs. The parameters of water composition are often stored in information systems in which the statistical treatment can be performed automatically by implemented statistical software, and then operators can be alerted when some outlaying samples are identified. Combination of both procedures of MDs calculations enables more reliable detection of unusual events.

Conclusion

The treated wastewater samples taken at the outlet of BWWTP during a year were evaluated by multivariate statistical analysis. The Mahalanobis distances calculated from the original 11 variables were evaluated according to their logMD magnitudes.

PCA was used for the reduction of data dimensionality from the original 11 variables to 4 principal components explaining 72% of the total data variance. The first principal component characterized the organic contamination of treated wastewater, and the second principal component characterized the nitrification processes in BWWTP. The third principal component indicated phosphorus-based compounds, and the fourth principal component presented the influence of effluent from the coking plant. Factor analysis and Ward's hierarchical clustering method confirmed the relationships between the original parameters revealed by PCA.

The MDs were calculated from the 4 PCs, and the logMD chart was plotted for all samples. The outlaying samples were identified, and their composition was discussed in relation to the treatment process. The samples represented by their PCs were also clustered by Ward's method into 3 clusters according to their compositions. On the whole, 16 outlaying samples (4.8% of all 334 samples) were found in the original and PC space and 7 of them were identified in both spaces.

It was shown that statistical processing of multidimensional data by means of the Mahalanobis distances calculated in the original and PC space could be complementary and should be useful for the evaluation of BWWTP performance. PCA enables data visualization in two- and three-dimensional spaces and understanding of the relationships between original variables. In contrast to the complexity and different magnitudes of the original parameters, the principle components visualized via their charts and scatter plots can provide a simple and effective tool for monitoring and evaluation of various technological processes.

Table 6 Composition of outlaying samples identified in original and PC space

N	NH_4^+ (mg/l)	BOD (mg/l)	COD (mg/l)	NO_3^- (mg/l)	PO_4^{3-}(mg/l)	CN^- (mg/l)	TN (mg/l)	TSS (mg/l)	TP (mg/l)	pH	TDS	Identification
21	4.91	9.9	46.0	47.3	0.633	0.126	14.1	15	1.1	7.7	864	Orig/PC
25	1.68	6.2	37.1	58.3	0.027	0.058	20.8	21	0.40	7.5	824	Orig/PC
95	0.095	4.1	32.7	49.0	0.367	0.143	17.0	7	0.94	8.0	804	Orig
96	0.789	3.6	28.1	14.8	1.40	0.074	6.9	3	0.69	7.6	364	PC
125	0.105	3.3	36.6	64.8	1.94	0.110	19.7	6	1.64	8.0	760	Orig
188	0.084	3.3	21.3	27.7	0.784	0.028	8.8	5	0.32	7.8	342	PC
200	1.71	6.1	28.6	47.5	1.44	0.117	47.8	8	0.70	8.4	758	Orig/PC
205	0.106	12.5	33.4	43.8	1.05	0.119	14.1	5	0.36	8.2	726	Orig
264	2.75	12.6	44.0	34.2	0.658	0.060	14.9	22	0.46	8.0	614	Orig/PC
270	0.166	13.8	50.6	28.1	2.53	0.196	14.6	11	1.14	8.0	842	Orig/PC
295	4.45	4.9	38.8	42.3	3.28	0.198	15.0	11	1.27	7.9	796	Orig/PC
296	5.08	4.7	37.7	41.0	3.61	0.167	17.7	9	1.29	7.8	806	Orig
298	6.05	5.1	39.1	35.4	1.18	0.132	19.1	8	0.49	7.8	812	Orig
329	0.948	10.7	75.2	42.5	0.861	0.097	24.0	22	0.92	7.7	730	Orig/PC
331	4.64	4.5	40.3	57.3	2.25	0.069	20.9	8	1.28	7.9	786	Orig
333	1.81	9.8	52.0	51.7	3.07	0.055	19.4	14	1.46	7.7	972	PC

Acknowledgements This work was financially supported by the National Feasibility Program I, Project LO1208 TEWEP.

References

Berthoex PM, Hunter WG, Pallesen L (1978) Monitoring sewage treatment plants: some quality control aspects. J Qual Technol 10:139–149

Capilla C (2009) Application and simulation study of the Hotelling's control charts to monitor a wastewater treatment process. Environ Eng Sci 26:333–341

Cobert CJ, Pan J (2002) Evaluating environmental performance using statistical control techniques. Eur J Oper Res 139:68–83

Everitt B (2001) Cluster analysis, 4th edn. Hodder Arnold, London

Hotelling H (1947) Multivariate quality control. In: Eisenhart C, Hastay MW, Wallis WA (eds) Techniques of statistical analysis. Mc-Graw-Hill, New York

Iglesias C, Sancho J, Piñeiro JI, Martínez J, Pastor JJ, Taboada J (2016) Shewhart-type control charts and functional data analysis for water quality analysis based on a global indicator. Desalin Water Treat 57:2669–2684

Jolliffe IT (2002) Principal component analysis, 2nd edn. Springer, New York

Kaiser HF (1960) The application of electronic computers to factor analysis. Educ Psychol Meas 20:141–151

Mac Nally R, Hart BT (1997) Use of CUSUM methods for water quality monitoring in storages. Environ Sci Technol 31:2114–2119

MacGregor JF, Kourti T (1995) Statistical process control of multivariate processes. Control Eng Pract 3:403–414

Maesschalck R, Jouan-Rimbaud D, Massart DL (2000) The Mahalanobis distance. Chemometr Intell Lab Syst 50:1–18

Mahalanobis PC (1936) On the generalised distance in statistics. Proc Natl Inst Sci India 2:49–55

Malinowski ER (1991) Factor analysis in chemistry, 2nd edn. Wiley, New York

Montgomery DC (1980) The economic design of control charts: a review and literature survey. J Qual Technol 12:75–89

Montgomery DC (1996) Introduction to statistical quality control, 3rd edn. Wiley, New York

Montgomery DC (2009) Introduction to statistical quality control, 6th edn. Wiley, Hoboken. ISBN 978-0-470-16992-6

Orssatto F, Vilasboas MA, Nagamine R, Uribe-Opazo MS (2014) Shewhart's control charts and process capability ratio applied to a sewage treatment station. Engenh Agríc Jaboticabal 34:770–779

Praus P (2005a) Water quality assessment using SVD-based principal component analysis of hydrological data. Water SA 31:417–422

Praus P (2005b) SVD-based principal component analysis of geochemical data. Cent Eur J Chem 3:731–741

Shewhart WA (1939) Statistical method from the viewpoint of quality control. The Graduate School, the U.S. Department of Agriculture, Washington, D.C.

Vega M, Pardo R, Barrado E, Debán L (1998) Assessment of seasonal and polluting effects on the quality of river water by explanatory data analysis. Water Res 32:3581–3592

Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:236–244